

Weekly Report

Done

1. 调研 document reading 相关可视化. 发现别人做的已经很多了 (主要思考的维度是: 对文本中何种数据进行可视化, 在文本中嵌入还是和文本并列, 单一文本还是多文本关联在可视化, 文本中数据的类型以及可视化方法)
2. 测试 VAE 在线采样上的结果, 不太理想, 换了一些激活函数也不佳, 似乎并不能学到一些想要的特征.
3. 了解、咨询一些图像中相似检测方法, 测试了一个相似计算方法, 感觉想要达到我们平时交互级的速度不可能.
 - a. 很早以前就有人做过了时间序列的可视化 query, 但是这个真的很简单, 限制了范围, 变化大小等等, 交互级毫无问题.
 - b. 也许针对某种特定的可视化, 比如热力图或者流图, 可以做这种相似性检测 (用户自己可视化结果框一个 pattern 系统推荐类似的区域), 并达到交互级, 但是看起来过于 trivial 贡献薄弱
4. 在一些可视化的结果上 (图片) 测试了 Grad-CAM, 主要是检测一个网络“看”到了图的哪一部分. 由于我们用的是现有模型 (所以他的 label 里不会有某种可视化, 只有猫狗降落伞等客观事物), 所以一个网络“看”到的都是可视化里面奇怪的东西. 这里应该就是类似一种 saliency 的问题, 其实现在有用 GAN 来生成 visual saliency.
5. 巫老师: 数据我们可以仍然用 twitter 和游戏数据; 题目 1 和彭老师聊聊应用相关的 2 自己想想能不能在时序的用户行为分析上做一个框架性方法, 并用游戏和 twitter 数据做 case.

To Do

1. 与彭老师聊一聊应用, 不管是行为演变还是网络上消息的散布等.
2. 测试一些方法. 最近看了一些 embedding 的技术, 他们不是简单的把序列类比为一个句子, 而是在 embedding 过程中加入了其余的信息.
3. 观望 PVis 的最终结果, 以待后续.

论文阅读

1. **InfoVis18:** Elastic Documents: Coupling Text and Tables through Contextual Visualizations for Enhanced Document Reading 将文本中表格数据提取出来并且和上下文关联的方法.
2. **InfoVis:** Dynamic query tools for time series data sets: Timebox widgets for interactive exploration 一种可视化查询方法, 里面有一些对单变量时间序列的常见的交互. 文章比较老.

3. Contextifier, NewsViews 都是一些多文本提取关键信息的可视化, 重温了一下. 还有 Temporal Summary Image 提取关键信息并在可视化上 annotation 的方法.
4. 一些 image retrieval, classification, detection 论文的泛读.
5. **NIPS 2017: Inductive Representation Learning on Large Graphs.** 一种名为 GraphSAGE 的方法, 直白的讲, 他不是简单的生成 embedding 结果, 而是导出如何 embedding 的方式. 比如对于动态网, 如果有后面新的点进来就可以很快的利用这个 embedding 得到比如分类结果等. 利用的是一种先聚合再挖掘信息.
 - a. 他们开源了代码但是要 10w+点才好用(因为有一个采样)所以我们也不便尝试.